

SOUND SYNTHESIZER

Patent number: JP56051795
Publication date: 1981-05-09
Inventor: SAGAYAMA SHIGEKI; ITAKURA FUMITADA
Applicant: NIPPON TELEGRAPH & TELEPHONE
Classification:
- international: G10L1/00
- european:
Application number: JP19790128364 19791003
Priority number(s): JP19790128364 19791003

Abstract not available for JP56051795

Data supplied from the *esp@cenet* database – Worldwide

⑨ 日本国特許庁 (JP) ⑩ 特許出願公開
⑪ 公開特許公報 (A) 昭56—51795

⑫ Int. Cl.³
G 10 L 1/00

識別記号 庁内整理番号
7350—5D

⑬ 公開 昭和56年(1981)5月9日
発明の数 1
審査請求 有

(全 6 頁)

④ 音声合成装置

⑪ 特願 昭54—128364
⑫ 出願 昭54(1979)10月3日
⑬ 発明者 嵐峨山茂樹
武蔵野市緑町3丁目9番11号日
本電信電話公社武蔵野電気通信

研究所内
⑭ 発明者 板倉文忠
武蔵野市緑町3丁目9番11号日
本電信電話公社
研究所内
⑮ 出願人 日本電信電話公社
⑯ 代理人 弁理士 草野卓

明細書

1 発明の名称 音声合成装置

2 特許請求の範囲

(1) 合成しようとする音声に応じた各正弦波成分の周波数、振幅及びピッチ周期を数ミリ秒ないし数十ミリ秒ごとに指定する手段と、上記指定されたそれぞれの周波数及び振幅を持つ複数の正弦波を発生する正弦波発生手段と、その正弦波の発生を上記指定されたピッチ周期ごとに位相を初期化する初期化手段と、それ等初期化された正弦波の和を合成して音声信号を得る合成手段とを具備する音声合成装置。

3 発明の詳細な説明

この発明は合成音声を発生する装置に関するものである。

音声合成装置は各種の通信機器、例えば音声応答装置、音声符号化復号化装置、音声出力つきの通信端末装置などや民生機器例えば教育機器、各種家庭電気器具などその他の人間の声を発生する各種の応用分野における音声合成の利用が考えられる。

られる。

従来の音声合成の方式として最も有力なものは PARCOR 形音声合成方式 (特許第876024号「音声合成装置」) あるいは LSP 形音声合成方式 (特許昭54—022318「音声合成装置」) などである。特に前者の方式は最近 LSI 化され広く応用される傾向が見えており、しかしこれらの従来の音声合成の方式には次のような欠点があつた。

例えば PARCOR 形音声合成方式は音声のスペクトル形状を PARCOR 係数と呼ばれる 8~12 個程度のパラメータにより表現するのであるが、これらのパラメータは音声現象との対応がつけにくく扱いが面倒であり、例えば規則による合成 (ルール合成) の場合この性質が障害となる。また量子化に関する性質が複雑で音声情報の圧縮率について問題を持つ。このために前に LSP 音声合成方式が発明されたのであるが、この方式はいまだ実用化の価格が高い。

いずれにせよこれら的方式は増進ループを持つ

(1)

(2)

離散フィルタ (recursive filter) を用いるため、発振や不安定性などの問題を内在している。このため例えば上記の音声合成方式をアナログ演算回路により実現することは要求される安定性や精度などの点からかなり難しく、一般にデジタル回路によつて実現する方が有利と考えられている。しかしデジタル回路を用いて実現する場合も演算時間が比較的長くする必要があつたり、パラメータの量子化に細心の注意を必要としたり種々の問題があつた。またそのため装置のハードウェア化も一般に複雑となり価格が高くなつてゐた。

この発明はこれらの欠點を根本的になくし音声合成フィルタを必要とせず、従つて比較的簡単なハードウェアで実現でき、またパラメータの量子化に関しても比較的粗くさえ演算時間を短かくすることができ、また合成音の品質も十分实用に耐え得、広い適用分野へ通用することができる音声合成装置を提供するものである。

まずこの発明による音声合成の原理を説明する。

(3)

離れ $t = 0, 1, \dots, 2^n - 1$ について等しくなるよう $\alpha_0, \dots, \alpha_n, A_1, \dots, A_n$ を決めることがで、これらを音声のフレームごとの分析結果とするものである。これは一種の音声スペクトル符号化と考えてよい。上述の文献では解の存在 ($0 \leq \alpha_i \leq 1, \alpha_i > 0, i = 1, \dots, n$) の条件や具体的な能率的な解法についても詳しく述べている。

この発明の音声合成の原理はパラメータ ($\alpha_0, \dots, \alpha_n, A_1, \dots, A_n$) を与えて (1) 式により合成音声 (x_t) を各時間 t についてつぎつぎと計算するものである。しかしこれだけでは音声としては聞こえない。なぜならば音声は有声音と無声音に分類されるが、有声音（母音や子音など）は周期的な信号であるが（この周期をビッチと言う）、上記のモデルにはその周周期性がない。つまりビッチの周波がないからである。そのためビッチ周期と共に (1) 式における時間 t をゼロにリセットし、つまり位相初期化する。これがこの発明のキーポイントの一つである。無声音についてはビッチは存

音声信号は短時間内ではほぼ定常と見なせる。前述の PARCO B 形音声合成方式においても BPF 形音声合成方式においても音声信号は短時間内では定常、即ちスペクトルが一定形状と見なして音声合成を行いそれぞれ成功している。この発明も同様に既非 S ないし板十回の短時間（フレームと呼ぶ）の音声を定常と見なし既往の正弦波の和で表現するものである。即ち離散的時間表現をすれば音声信号端系列 $\{x_t\}$ を

$$x_t = A_1 \sin \omega_1 t + \dots + A_n \sin \omega_n t \quad (1)$$

のようくモデル化する。但し ω は離散的な時間を表わす振幅、 ω は正弦波成分の角周波数（通常 4 ~ 6 倍程度）、 A_i は第 i 正弦波成分の角周波数（0.5 ~ 1.5 Hz）、 A_1 は第 1 正弦波成分の振幅である。このモデルを複合正弦波モデルと呼ぶ。このモデルによる音声分析法については日本音響学会音声研究会資料 87.9-06 (1979年 6月) 赤坂山・改定「複合正弦波モデルによる音声の分析」で詳しく述べられてある。この分析法は (1) 式（モデル）の自己相關関数と実際の音声の自己相關関数とが

(4)

近しないが (1) 式のままではスペクトルが日本の標準スペクトルとなつて音声信号の実感とは異なるためランダムなビッチ周期を与えてランダムに位相初期化することによりスペクトルを広げる。

以上の原理により構成した音声合成装置の例を第 1 図に示す。電圧制御正弦波発振器 1, 1 ~ 1, が設けられそれぞれ端子 2, 1 ~ 2, を通じて与えられる角周波数 $\omega_1 \sim \omega_n$ に対応する電圧が電圧制御正弦波発振器 1, 1 ~ 1, に入力され、振幅 1, 角周波数 $\omega_1 \sim \omega_n$ の正弦波をそれぞれ出力する。これら発振器 1, 1 ~ 1, はリセプト端子を持ちこれにパルスを与えると正弦波出力の位相はゼロから再スタートする。ビッチ周期（あるいは周波数）に対応する電圧が端子 1, 1 を通じて電圧制御パルス発生器 1, 2 に与えられ、このパルス発生器 1, 2 からそのビッチ周期ごとにパルスが発生される。このビッチ周期パルスにより電圧制御正弦波発振器 1, 1 ~ 1, はリセプトされる。電圧制御正弦波発振器 1, 1 ~ 1, の出力は電圧制御可変利得増幅器 3, 3 に供給されて振幅 1 の正弦波に端子 4, 4, エリの

(5)

(6)

振幅 $A_1 \sim A_8$ がそれぞれ乗算される。これ等乗算出力は加算器 1-3 で合算されて出力端子 1-4 へ供給される。

動作は上述の振幅に値すき合成分しようとする音声と対応して数回 B ないし數十回 S ごとに端子 2, ~ 2, 4, ~ 4, 及び 1-1 に与える角周波数、振幅、ピッチ周期にそれに対応した入力電圧を変化させてゆくとそれに対応した合成音声が出力から得られる。

第 1 図は各正弦波成分の振幅を面接与えてそれらの和を出力としたものであるが、一般に音声のダイナミクレンジは広いため、各正弦波成分の振幅 ($A_1 \sim A_8$) も大きなダイナミクレンジを持つ。実験的検討によれば(1)式を

$$x_t = A (a_1 \sin \omega_1 t + \dots + a_n \sin \omega_n t) \quad (1)$$

但し $a_1^2 + \dots + a_n^2 = 1, 0 < a_i < 1 (i=1 \dots n)$ のよう倍音を放わし ($A_1 \sim A_8$) のかわりに ($A, a_1 \sim a_n$) を与えた方が音声合成パラメータの情報圧縮符号化に有利である。今後 ($a_1 \sim a_n$) を各正弦波成分の係数、A を利得と呼ぶこ

(7)

(8)

第 3 図にその例を示し電圧調節ペルス発生器 1-2 のペルス出力が包絡線発生器 1-7 のトライガ入力に与えられ、このペルスに同期して振幅が端子 1-6 の利得電圧に比例した包絡線信号が発生される。包絡線信号としては具体的には指數関数やレイムドコサイン関数 (($e^{j\theta} + 1$) / 2) が用いられる。この包絡線信号が乗算器 1-5 へ供給される。

以上に述べた装置においては角周波数 (ω_1)、係数 (a_1)、利得 A 及びピッチ周期などの制御入力は数回 B ないし數十回 S ごとに変化させてゆくことを前提にしていたが、これらの入力は階級状に変化させるより滑らかに変化させた方が良好な合成音声が得られることが実験的にわかつている。そのため各正弦波成分の周波数の制御電圧及び振幅 (あるいは係数と利得) の制御電圧をなめらかに変化させるためのフィルタ (一種の低域通過フィルタ) に通してから与えるのがよい。第 4 図はその例を示し平滑化フィルタ 2-11, 2-21 (l = 1, 2, ..., L) により階級状に変化する端子 2, 4 の角周波数 (ω_1)、係数 (a_1) の制御電圧をそれぞれ

(9)

にする。各正弦波成分の周波数と係数、利得及びピッチ周期を与えて音声合成をする場合のこの簡単の実例を第 2 図に示す。ここではロードの場合について示している。第 2 図中第 1 図と対応する部分には同一符号を付けてある。この例では端子 4, ~ 4, からそれぞれ係数 $a_1 \sim a_8$ を示す信号を可変利得増幅器 3, ~ 3, へ供給する。加算器 1-3 の出力はレベル制御器 1-5 で端子 1-6 よりの利得を示す信号と乗算されて出力される。可変利得増幅器 3, ~ 3, の代りに乘算器あるいは可変減算器減算器を用いてもよい。レベル制御器 1-5 としては乘算器あるいは可変利得増幅器または可変減算器減算器を用いることができる。

以上述べた装置においてはピッチ周期ビート位相を初期化しているが、本質的には正弦波の和のみで合成音声を作り出そうとしていた。しかし実際の音声信号は数本の線スペクトルで構成されるものではなく、一般に複数スペクトルである。そこで線スペクトルの幅を広げるためにピッチ周期に同期した包絡線パターンを乗じることが好ましい。

滑らかにする。

次に以上説明した装置の各部の詳細例を述べる。第 5 図は発振周波数が電圧により制御されリセットすれば位相が 0° に初期化される電圧制御正弦波発振器の例である。端子 2-1 の電圧によって電流が制御される電流源 2-3, 2-4 が設けられ、これら電流源 2-3, 2-4 は電流スイッチ 2-5 により切替えられてコンデンサ 2-6 の一端及び電圧比較器 2-7, 2-8 に接続される。コンデンサ 2-6 上の入力電圧が基準電圧 V_r より高いと比較器 2-7 の出力が論理 1 になり、入力電圧が基準電圧 $-V_r$ より低いと比較器 2-8 の出力が論理 1 になり、その他は論理 0 を出力している。比較器 2-7 の出力により K8 フリップフロップ 2-9 がリセットされその出力は電流スイッチ 2-5 を制御する。端子 2-1 の状態が 1 (セント) のときスイッチ 2-5 は電流源 2-3 の側に接続され、コンデンサ 2-6 は充電される。フリップフロップ 2-9 の状態が 0 (リセット) のときスイッチ 2-5 は電流源 2-4 の側に接続されコンデンサ 2-6 は放電される。端子

(10)

3-1に位相初期化入力パルスが加えられると位相初期化のためのスイッチ3-2が閉じてコンデンサ3-6が充電されるとともにR8フリップフロップ3-9がセットされる。このようにしてコンデンサ3-6の電位は三角波となりその周波数は電源23、24に対する端子21の制御電圧により変化させられる。また端子3-1より位相初期化入力にパルスを加えるとR8フリップフロップ3-9がセットされるとともにスイッチ3-2が閉じられてコンデンサ3-6が充電され、前記三角波の位相は0°に初期化される。コンデンサ3-6の三角波電圧は正弦波変換器3-3により正弦波へ変換されて端子3-4へ送出される。正弦波変換器3-3としては通常折線近似が用いられる。以上に述べた電圧制御発振器の市販品の好例は米国インターシル社製の電圧制御発振器IC(型番ICL8038)である。これには上記の位相初期化入力がないが、その他は第5図に示したものとは同様の構成となつており、実験若干の付加回路によりこの発明の装置において利用できる。

(11)

ータ(balanced modulator)回路や各種のエキスパンダーやコンプレッサ(例えばドルピー録音低減方式やdbxノイズリダクションシステムにおいて使用されているものなど)が利用できる。

電圧制御パルス発生器1-2は周波変化範囲は80~500Hz程度と比較的広いので各種の実現法が利用できる。簡単なものとしては電圧で操作できる電源源とコンデンサと充電回路を組合せたものがある。その一例を第7図に示す。端子1-1及び電源端子4-5との間の制御電圧がトランジスタ4-6のベースに与えられてそのコレクタ電流が制御され、そのコレクタ電流は抵抗4-7を通じてコンデンサ4-8へ供給される。抵抗4-7の内側はダイヤ(米国シグネナイス社製NBS56)4-9のトリガ端子、充電端子及びしきい値端子に接続され、このダイヤ4-9よりパルスが端子5-1へ送出される。このほか各種のパルス変換回路も利用できる。

周波電圧の動きを平滑化フィルタは簡単なものとしては第8図に示すC-R1次フィルタが利用でき

る。計算あるいは可変利得増幅器あるいは可変減衰率増幅器を具体的に実現するには様々な方法がある。市販されている各種の乗算・除算用計算機として利用できる。また第6図に示すような簡易な回路によつても利得を電圧または電流によつて制御できる増幅器が実現できる。図中Aは入力端子3-5を抵抗3-6を通じてD端子T3-7のドレインに接続しD端子T3-7のソースを絶縁しゲートに端子4から制御電圧を与え、D端子T3-7のドレイン電圧を出力端子3-8へ送出する。第6図Bは入力端子3-5とD端子T3-7のドレインを計算増幅器3-9の非反転入力端及び反転入力端にそれぞれ接続し計算増幅器3-9の出力端子と反転入力端との間に抵抗4-1が接続される。第6図Cは第6図AのFET3-7の代りに発光ダイオード4-2及び光導電端子4-3のフォトカプラ4-4が用いられる。第6図Dは第6図Cの抵抗3-6とフォトカプラ4-4とを入れ替えたものである。第6図Eは第6図BのFET3-7の代りにフォトカプラ4-4を用いた場合である。この他にもバランスストセレクタ

(12)

きる。包絡線の発生は指紋開数の場合は非常に簡単であり、第9図に示すようにパルス発生器1-2の出力パルスによりスイッチ5-2が瞬時に閉じられ、その瞬間端子1-6の電圧にコンデンサ5-3が充電されスイッチ5-2が閉じるとコンデンサ5-3の電荷は抵抗5-4を通じて放電される。コンデンサ5-3の電圧は端子5-5に包絡線出力5-6として出力される。

以上説明したようにこの発明は従来の音声合成装置に比較して極めて簡単な構成を持ちハードウェアも容易に実現できる。しかも合成音声の品質は比較的良好であるという特徴を持つ。また音声合成フィルタをまつたく使用せず、直接的に波形を合成するものであるため帰還ループの安定性や発振防止の問題を本質的に持たない。このためかなり難な取扱いにも耐え得る。従つてこの発明の音声合成装置は安価な音声合成システムとして有用であり、その応用分野としては歌や空船などのアナウンスマシン、銀行のキャッシュディスペンサー、電話の音楽内、外音楽教育機器、音声出力つき

(13)

(14)

のほん証機械、音声出力つきのコンピュータ端末、その他各種の音声出力つきの通信機器や民生機器など多岐にわたることができる。なお上述ではこの発明をアナログ信号処理により構成したが、デジタル処理により同様に構成することができる。この場合は各同一の機能部分は各正弦波の系統について専分割多路利用することができる。

4. 製品の簡単な説明

第1図はこの発明による音声合成装置の実施例($n=5$ としている)を示すブロック図、第2図はこの発明の他の実施例を示すブロック図、第3図は出力の包絡線をピッチ同期で変化させる例を示すブロック図、第4図は周波電圧を滑らかに変化させる例を示す図、第5図は位相初期化可能な電圧制御正弦波発振器の例を示す回路図、第6図は電圧制御可変利得増幅器及び可変減衰率減振器の各種例を示す回路図、第7図は電圧制御可変周波数バルス発生器の例を示す回路図、第8図は制御電圧の平滑化器の例を示す回路図、第9図は包絡線発生器の例を示す回路図である。

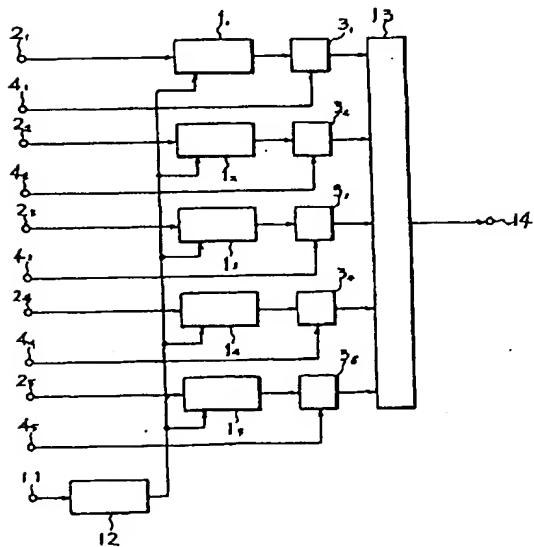
(15)

(16)

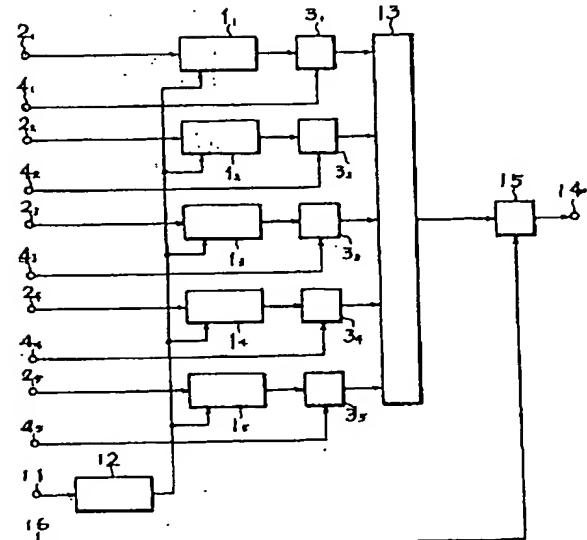
特許出願人 日本電信電話公社

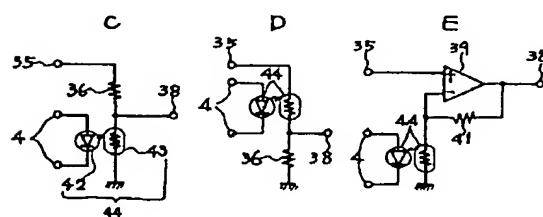
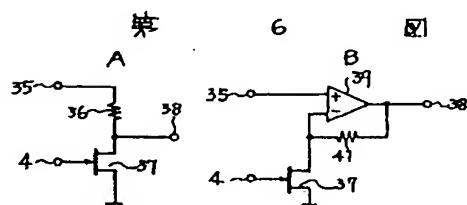
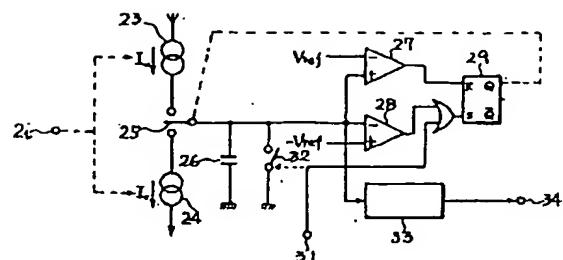
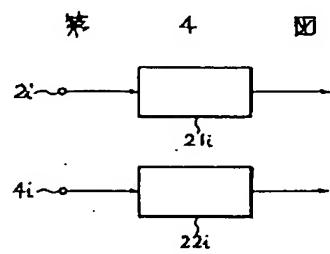
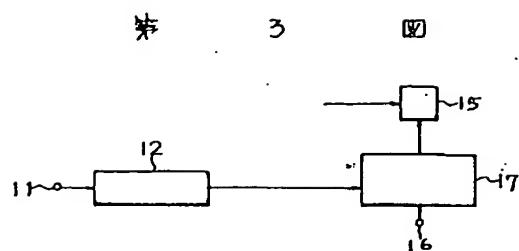
代理人 単野 单

第一 図



第二 図





第 8 図

